Question 1a/1b

1a: Define

$$a_{i,j} = \frac{\exp\{\theta_{i,j,n}\}}{\sum_{i=1}^{n} \exp\{\theta_{i,j,n}\}}$$

1c: set $\theta_{1,1,2}$ and $\theta_{2,2,2}$ to large values, and all other θ values to 0.
It follows that $a_{1,1}$ and $a_{2,2}$ are then close to 1. It follows that

$$c^{(1)} \approx u^{(1)}$$

and

$$c^{(2)} \approx u^{(2)}$$

Question 2a

- Input: a source-language sentence $x_1 \dots x_n$
- Initialization: INIT(BEAM)
- For $k = 1 \dots m$:
  - Foreach $(y_1 \dots y_{k-1}, \text{score}) \in \text{BEAM}(k-1)$
    - Calculate $\log p(y|y_1 \dots y_{k-1}, x_1 \dots x_n)$ for each y in the vocabulary, using the computational graph
    - For each y in the vocabulary,

      $$\text{ADD}(y_1 \dots y_{k-1} y, \text{score}')$$

      where

      $$\text{score'} = \text{score} + logp(y|y_1 \dots y_{k-1}, x_1 \dots x_n)$$
- Return ARGMAX(BEAM)